

Biometry of the Characteristics of Cajazeira (*Spondias mombin* L.) Stone in Northeast Brazil

Janilson Pinheiro de Assis¹, Roberto Pequeno de Sousa¹, Paulo César Ferreira Linhares¹,
Eudes de Almeida Cardoso¹, José Aluísio de Araújo Paula², Lunara de Sousa Alves³,
Maria Francisca Soares Pereira¹, Cydianne Cavalcante da Silva³,
Glenda Soares de Lira Rosado Nogueira¹ & Misael Bruno de Araujo Silva³

¹ Federal Rural Semi-arid University, Mossoró, RN, Brazil

² Brazilian Service to Support at Micro and Small Enterprises of Rio Grande do Norte (SEBRAE-RN), Mossoró, RN, Brazil

³ Federal University of Campina Grande, PB, Brazil

Correspondence: Janilson Pinheiro de Assis, Jitirana Research Group, Department of Agronomic and Forestry Sciences, Federal Rural Semi-arid University, Mossoró, RN, Brazil. E-mail: janilson@ufersa.edu.br

Received: May 21, 2019

Accepted: November 3, 2019

Online Published: December 15, 2019

doi:10.5539/jas.v12n1p58

URL: <https://doi.org/10.5539/jas.v12n1p58>

Abstract

The cajazeira (*Spondias mombin* L.) is a stone tree belonging to the Anacardiaceae family is very common in the caatinga ecoregion of northeastern of Brazil. This study assessed the biometry of cajazeira stones. The stones were collected in 2019 within an area of native vegetation in the campus of the Federal Rural University of the Semi-arid Region (UFERSA), Mossoró-RN. They were then taken to the Plant Science Laboratory to measure the following characteristics: length, width, length/width ratio, stones thickness, and weight. The variables length, width, length/width ratio, thickness, and weight of the stones had a low range of variation and low values for relative variation, and therefore high coefficients of Pearson of correlation. The data of length, width, and thickness showed a moderate degree of symmetry and a symmetrical and mesocurtic distribution, while length/width ratio and stones weight had almost symmetrical and mesocurtic distributions. The variables length and width and width and thickness showed significant positive linear correlations. The biometric data of caju stones fitted in to the approximate normal distribution of probability.

Keywords: *Spondias mombin* L., stone species, descriptive statistics, inference

1. Introduction

The cajazeira (*Spondias mombin* L.) it is a tree belonging to the family Anacardiaceae. This species has hermaphrodite, staminate and pistillate flowers, with an apparent occurrence of protandry, which favors cross-pollination and genetic variability in orchards (Sacramento & Souza, 2000). Is originated of the tropical region of the American continent. In the Brazil, the plant occurs mainly in the north and northeastern regions, where its fruits receive different denominations, such as “cajá”, “cajá verde”, “cajá-mirim”, or “taperebá” (Bosco et al., 2000).

Besides the commercial value of its pulp, according to Sacramento and Souza (2000), the wood of the cajazeiras used in carpentry. However, the species is still very little cultivated on a commercial scale, being considered a plant in domestication and extractive exploitation. Even so, it has growing participation in the agribusiness of the Northeast region, mainly by the commercialization of its stones for pulp processing.

Commercialization in the Northeast region of Brazil is carried out in street markets, at the margins of roads near the production units, and in the local pulp processing industries. In the absence of commercial orchards, agroindustries of cajazeiras tones depend on extractive production, which is seasonal and insufficient for the operation of pulp agroindustries (Souza et al., 2005).

The stones of the mombin, also known as cajazeira, are indiscriminate, perfumed drupe, with meaty yellow mesocarp, bittersweet flavor, containing carotenoids, sugars, vitamins A and C (Sacramento & Souza, 2009).

The structure for sexual propagation of mombin comprises the endocarp containing stones, also known as the kernel, which is lignified and surrounded by spongy fibers (Carvalho et al., 1998). According to Sacramento and Souza (2009), the stone contains two to five loci, with none up to five stones. Pereira (1996) working with three samples of 100 endocarps, observed up to eight loci and four stones per endocarp.

The size and length of the cajazeira endocarp also indicate the increase of mass, number of locules and number of stones per endocarp (Quadros et al., 2015). Studies involving morphological analyzes of stones and stones can help to understand the germination process and the characterization of vigor and viability (Matus & Lopes, 2007). On the other hand, biometric analyzes are an essential tool to detect genetic variability within and between populations and to define the relationships between this variability and environmental factors, thus contributing to genetic improvement programs (Gusmão et al., 2006).

The size and characteristics of the stones are of great importance for the study of a plant species. It is a fundamental parameter to understand stone dispersal and establishment (Fenner, 1993). It is also used to differentiate pioneer and non-pioneer species in tropical forests (Baskin & Baskin, 1998).

The biometric characterization of stones and stones can provide significant subsidies for the differentiation of species of the same genus. For example, Carpanezzi and Marques (1981) showed that the weight of the stones of *Hymenaea courbaril* is almost twice the weight of *Hymenaea parvifolia* stones. In most cases, for shrub and tree species, there is antagonism between stone size and number of stones per stone, as observed by Carvalho et al. (1998).

Stones play an essential role in human health, contributing to the supply of calories, minerals, vitamins, fiber, and water (Feitosa, 2007). Despite the real and potential importance of the stones of the cajazeira to the Northeast, there is little technical information about it. The specialized literature is dispersed and little comprehensive, the results are limited in scope due to the restricted dissemination (Mendonça et al., 2008).

The scientific research responses regarding the descriptive measures of location, variability or scale, asymmetry and kurtosis can serve as the basis for future studies of descriptive analysis and statistical inference, for the comparison of different environments, genetic studies and genetic improvement, criteria used for the grouping of experiments in joint analysis, in stability analysis of cultivars, in multivariate analysis, as well as in the construction of known variance components, for simulation and modeling studies applied to agriculture.

In view of the above, the objective was to evaluate the biometric variables of the cajazeira (*Spondias mombin* L.) stone in the Northeast of Brazil.

2. Materials and Methods

The study was carried out in Mossoró, Rio Grande do Norte-RN, Brazil, in 2019, whose geographic coordinates are: 5°11'S and 37°20'W, with 18 m altitude, annual average temperature around 27.5 °C and relative humidity of 68.9% (Carmo Filho et al., 1991). According to Kottek et al. (2006) and the classification of Köppen, the local climate is BSw^h, dry and very hot, the dry season being normally from June to January, and a rainy season being from February to May. The average annual rainfall is 673.9 mm and the average relative humidity is 68.9%.

A total of 300 caja stones were collected in an area of native vegetation, located within the campus of the Federal Rural University of the Semi-arid Region (UFERSA) in January 2018 and were taken to the plant science laboratory, where the following characteristics were evaluated: a) morphological characterization of the stone, determining the length and width of 300 well-developed stones, with the aid of a caliper; (b) stone thickness in millimeters; and (c) stone weight in grams.

Descriptive and graphic analyzes were performed using the software package R, version 3.5.1. (2019).

3. Results and Discussion

To measure and compare important quantitative aspects of the values distributions of variables length, width, length/width, thickness and stones weight, the present work was based on the specialized biometric literature (Ferreira, 2005; Spiegel & Stephens, 2009; Oliveira et al., 2009; Bussab & Morettin, 2010; Zar, 2010; Claudio & Stein, 2011; Cecon et al., 2012). In this way, we applied the exploratory data analysis using frequency distributions of series, statistical series or heterogeneous series, box plot graphs, as well as the statistical estimators of the variables under study, which are the typical or summary measures called position or location, dispersion, variation or scale measurements, measures of skewness or bias.

Therefore, we used the known descriptive or deductive statistics, such as the mean, coefficient of variation, skewness, kurtosis, the quartiles, the interquartile range, coefficient of Pearson's for simple linear correlation, as

well as the application of statistical inference techniques such as hypothesis tests using the T-test and the Z-test at a significance level of 5% of probability, based on Student's t distribution and Normal distribution, respectively, in the estimation of 95% probability confidence intervals (Tables 1-7, Figures 1-11).

Table 1. Descriptive and inductive statistical analysis of 300 sample units for the variables length, width, length/width ratio, thickness (mm) and weight in grams (g) of caja stones (*Spondias mombin* L.)

| Sample or estimator statistics | Length | Width | Length/width ratio | Thickness | Weight |
|---|---|---|---|---|---|
| Sample Size (number of fruits) | 300 | 300 | 300 | 300 | 300 |
| Minimum value | 24.2 | 12.1 | 1.5 | 12.2 | 1.09 |
| Maximum value | 34.6 | 20.2 | 2.3 | 21.7 | 3.4 |
| Total Range | 10.4 | 8.2 | 0.86 | 9.5 | 2.3 |
| Arithmetic Average | 30.2 | 16.0 | 1.9 | 16.5 | 1.8 |
| Median or Second Quartile | 30.3 | 16.0 | 1.9 | 16.5 | 1.8 |
| Variance | 3.2 | 2.2 | 0.02 | 2.3 | 0.14 |
| First quartile | 29.1 | 15.2 | 1.8 | 15.6 | 1.6 |
| Third quartile | 31.5 | 16.9 | 2.0 | 17.4 | 2.1 |
| Standard deviation | 1.8 | 1.5 | 0.15 | 1.5 | 0.4 |
| Average standard error | 0.1 | 0.09 | 0.0088 | 0.09 | 0.02 |
| Coefficient of variation (%) | 5.9 | 9.3 | 8.03 | 9.11 | 20.47 |
| Skweness | -0.4 | -0.11 | 0.1081 | -0.03 | 1.00 |
| Kurtosis | 0.4 | 0.34 | 0.1598 | 0.71 | 1.77 |
| Interquartile Range (IR) | 2.4 | 1.75 | 0.2 | 1.87 | 0.48 |
| Z Test for the mean at 0.1% probability | 302*** | 178*** | 215*** | 183*** | 90*** |
| Confidence Interval for the average 95% probability | 30.0 to 30.2 | 15.7 to 16.2 | 1.73 to 1.92 | 16.3 to 16.7 | 1.76 to 1.84 |
| Fitting to Normal Distribution (D'Agostino-Pearson Test) | Valor p = 0.01 fit to Normal Distribution | Valor p = 0.73 fit to Normal Distribution | Valor p = 0.01 fit to Normal Distribution | Valor p = 0.01 fit to Normal Distribution | Valor p = 0.05 fit to Normal Distribution |

Table 2. Distribution of frequencies of the length of the stone of caja (*Spondias mombin* L.)

| Length (mm) | f _i | X _i | f % |
|---------------------|----------------|----------------|-------|
| 24.20 [-----) 25.50 | 5 | 26.15 | 1.67 |
| 25.50 [-----) 26.81 | 5 | 27.46 | 1.67 |
| 26.81 [-----) 28.11 | 27 | 28.76 | 9.00 |
| 28.11 [-----) 29.41 | 59 | 30.06 | 19.67 |
| 29.41 [-----) 30.71 | 86 | 31.36 | 28.67 |
| 30.71 [-----) 32.02 | 77 | 32.67 | 25.67 |
| 32.02 [-----) 33.32 | 32 | 33.97 | 10.67 |
| 33.32 [-----) 34.62 | 9 | 24.85 | 3.00 |
| Total | 300 | - | 100 |

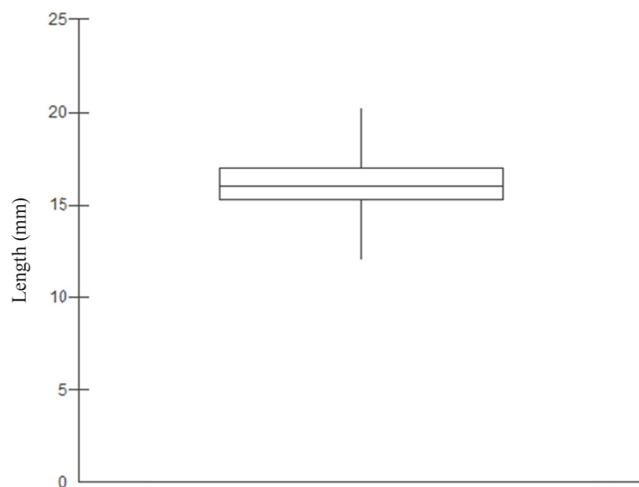


Figure 1. Boxplot with median and quartiles for the length of the caja stones (*Spondias mombin* L.)

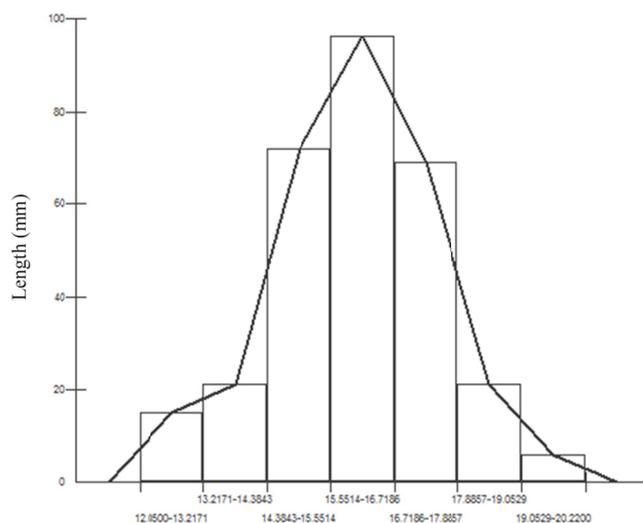


Figure 2. Histogram and polygon of frequencies for the length distribution of the caja stones (*Spondias mombin* L.)

Table 3. Frequency distribution of the width of caja stones (*Spondias mombin* L.)

| Width in mm | f_i | X_i | f % |
|----------------|-------|-------|-------|
| 12.05 — 13.22 | 15 | 12.63 | 5.00 |
| 13.22 — 14.38 | 21 | 13.80 | 7.00 |
| 14.38 — 15.55 | 72 | 14.97 | 24.00 |
| 15.55 — 16.72 | 96 | 16.14 | 32.00 |
| 16.72 — 17.89 | 69 | 17.30 | 23.00 |
| 17.89 — 19.05 | 21 | 18.47 | 7.00 |
| 19.05 — 20.22 | 6 | 19.64 | 2.00 |
| Total | 300 | - | 100 |

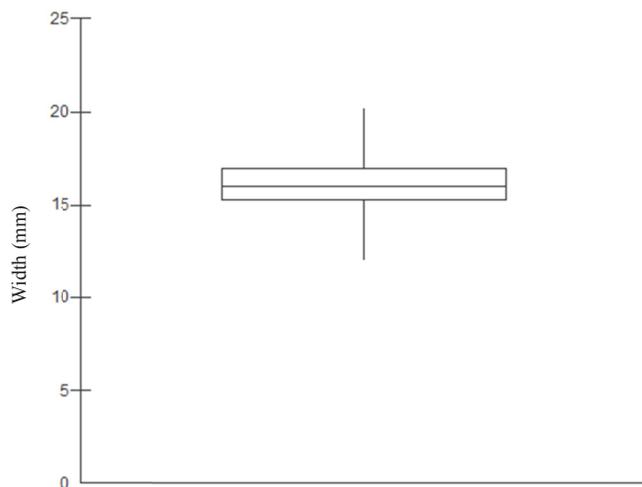


Figure 3. Box plot with median and quartiles for the width of the caja stones (*Spondias mombin* L.)

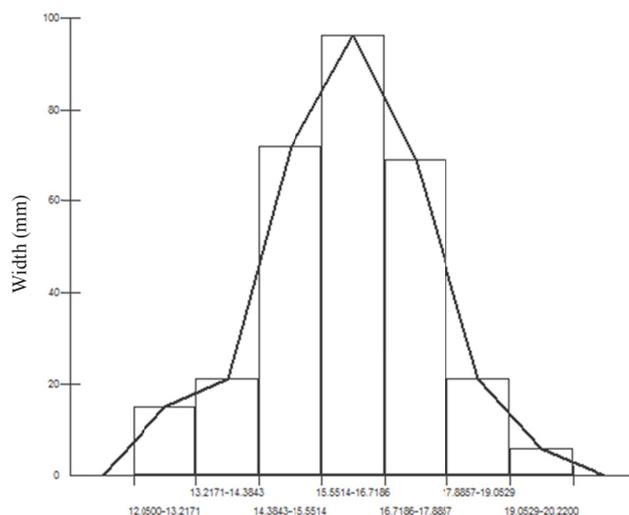


Figure 4. Histogram and polygon of frequencies representative of the width distribution in millimeters of the caja stones (*Spondias mombin*)

Table 4. Distribution of frequencies of the length/width ratio of caja stones (*Spondias mombin* L.)

| Length/Width ratio | f_i | X_i | $f\%$ |
|--------------------|------------|-------|------------|
| 1.0 — 1.1 | 3 | 1.1 | 1.00 |
| 1.1 — 1.2 | 1 | 1.2 | 0.33 |
| 1.2 — 1.3 | 17 | 1.3 | 5.67 |
| 1.3 — 1.4 | 104 | 1.4 | 34.67 |
| 1.4 — 1.5 | 137 | 1.5 | 45.67 |
| 1.5 — 1.6 | 33 | 1.6 | 11.00 |
| 1.6 — 1.7 | 3 | 1.7 | 1.00 |
| 1.7 — 1.8 | 2 | 1.8 | 0.67 |
| Total | 300 | - | 100 |

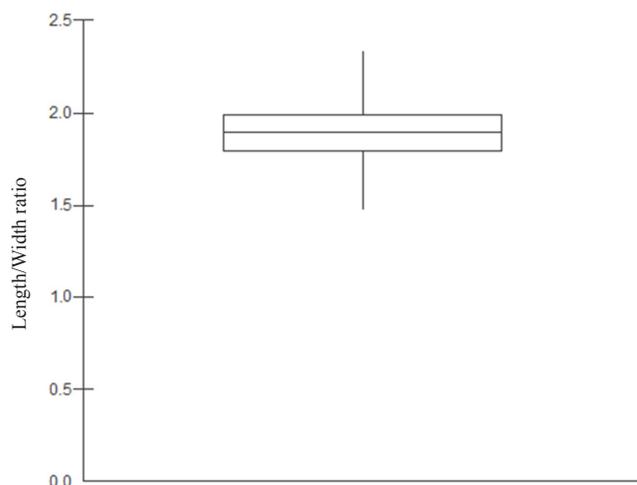


Figure 5. Boxplot with median and quartiles for the length/width ratio of caja stones (*Spondias mombin* L.)

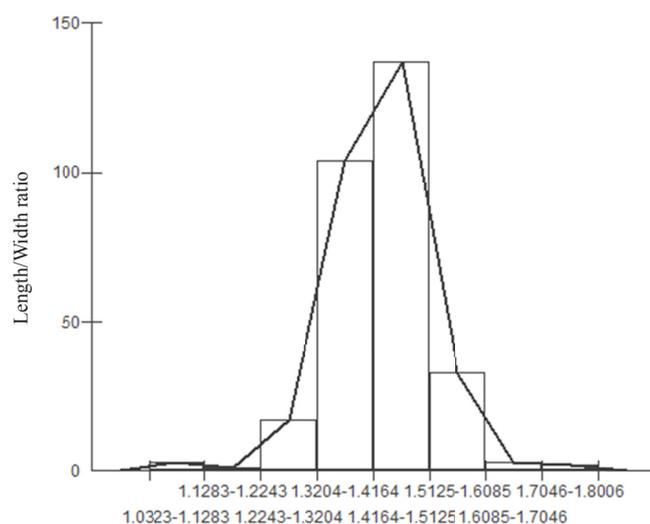


Figure 6. Histogram and polygon of frequencies for the distribution of length/width ratio of caja stones (*Spondias mombin* L.)

Table 5. Distribution of frequencies of the thickness of caja stones (*Spondias mombin* L.)

| (Thickness in mm) | f_i | X_i | f % |
|-------------------|-------|-------|-------|
| 23.6 — 25.4 | 16 | 24.5 | 5.33 |
| 25.4 — 27.2 | 67 | 26.3 | 22.33 |
| 27.2 — 28.9 | 135 | 28.1 | 45.00 |
| 28.9 — 30.8 | 66 | 29.9 | 22.00 |
| 30.8 — 32.6 | 12 | 31.7 | 4.00 |
| 32.6 — 34.4 | 3 | 33.5 | 1.00 |
| 34.4 — 36.2 | 1 | 35.3 | 0.33 |
| Total | 300 | - | 100 |

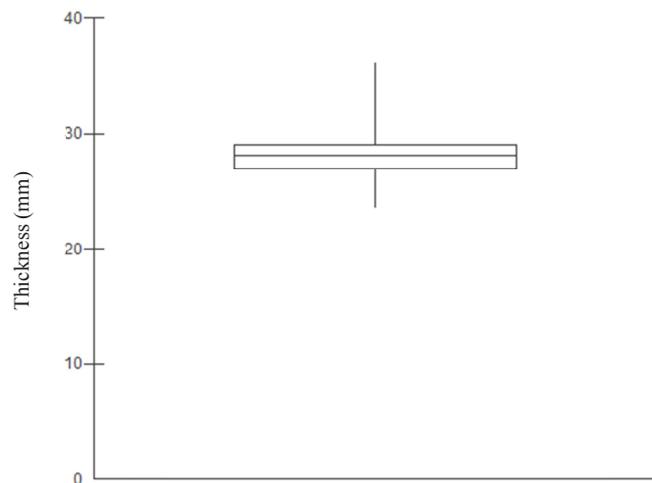


Figure 7. Boxplot graph with median and quartiles of the thickness of caja stones (*Spondias mombin* L.)

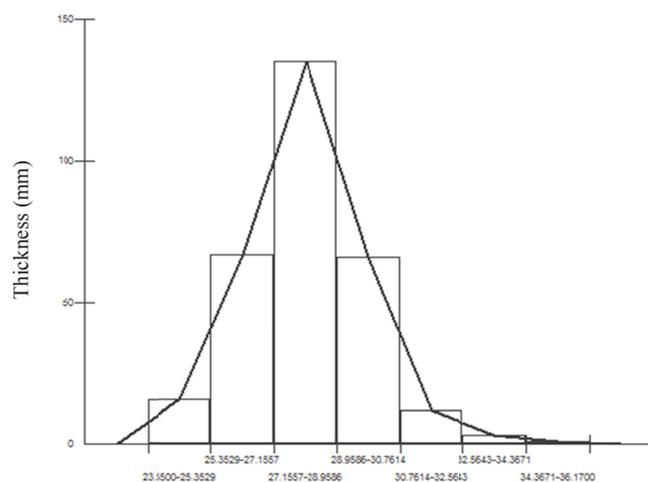


Figure 8. Histogram and polygon of frequencies for the thickness distribution of caja stones (*Spondias mombin* L.)

Table 6. Distribution of frequencies of the weight in grams of the caja stones (*Spondias mombin* L.)

| (Weight in g) | f_i | X_i | f % |
|---------------|-------|-------|-------|
| 11.0 — 13.0 | 13 | 12.0 | 4.33 |
| 13.0 — 15.0 | 47 | 14.0 | 15.67 |
| 15.0 — 17.0 | 81 | 16.0 | 27.00 |
| 17.0 — 19.0 | 88 | 18.0 | 29.33 |
| 19.0 — 21.0 | 48 | 20.0 | 16.00 |
| 21.0 — 23.0 | 17 | 22.0 | 5.67 |
| 23.0 — 25.0 | 2 | 24.0 | 0.67 |
| 25.0 — 27.0 | 4 | 26.0 | 1.33 |
| Total | 300 | - | 100 |

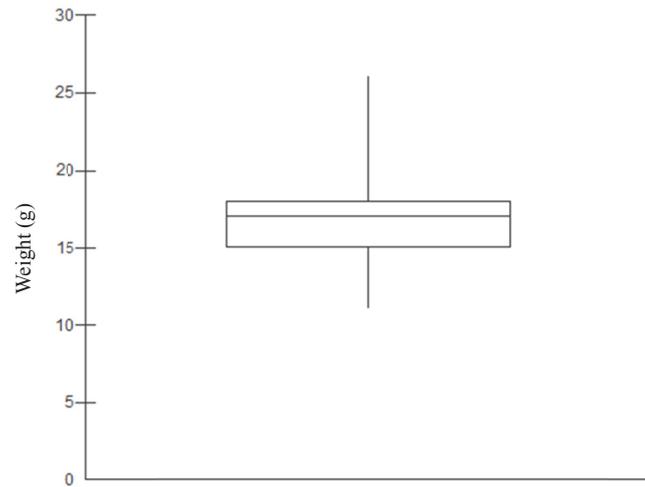


Figure 9. Boxplot with median and quartiles of the weight in grams of caja stones (*Spondias mombin* L.)

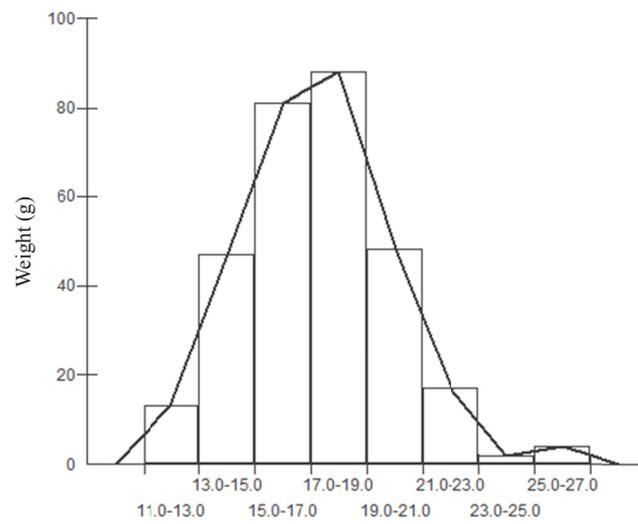


Figure 10. Histogram and polygon of frequencies for the weight distribution in grams of caja stones (*Spondias mombin* L.)

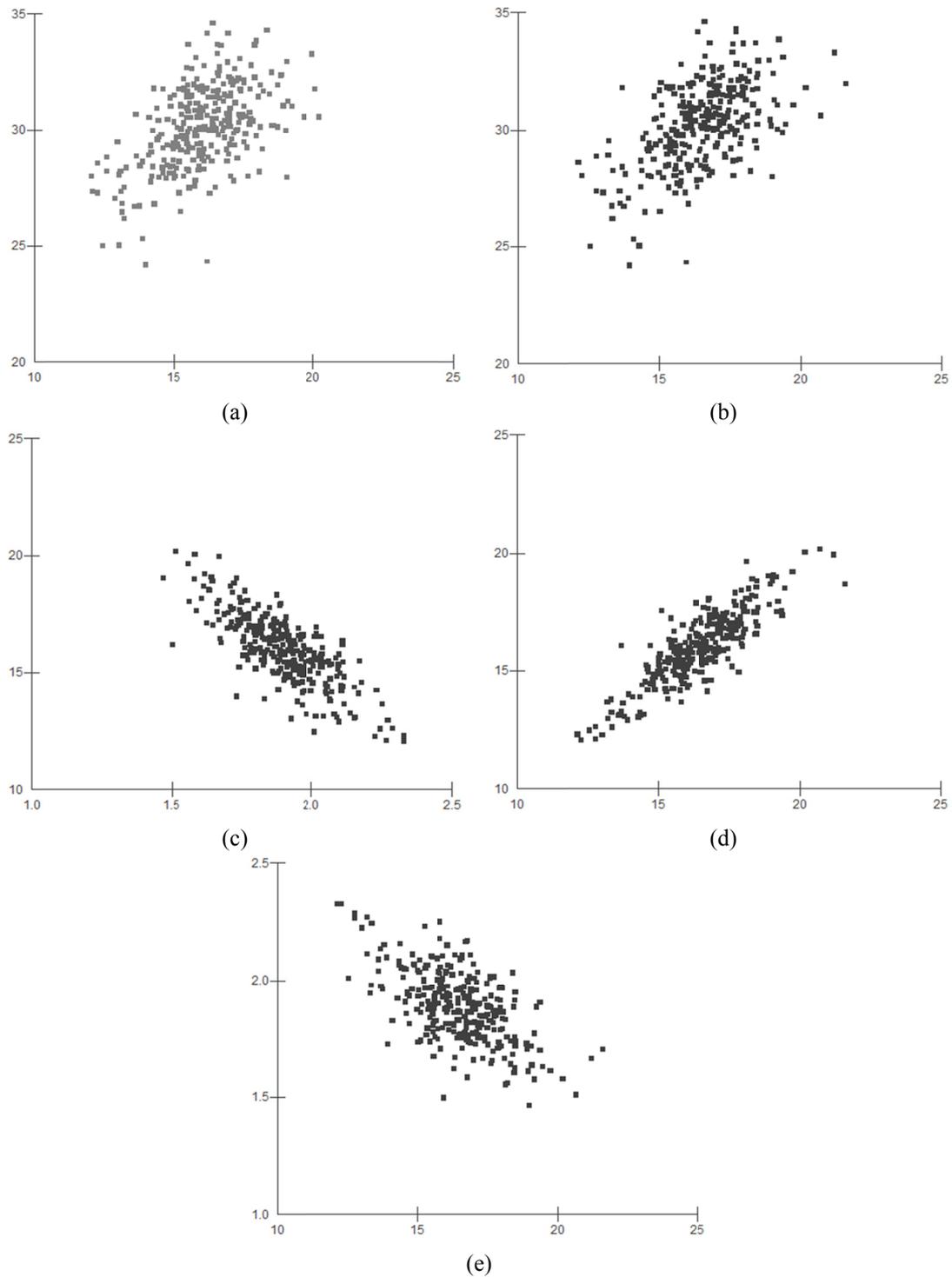


Figure 11. Dispersion plots of correlations between pairs of variables: length versus width (a), length versus length/width ratio (b), length versus thickness (c) and width versus thickness (d) (e) thickness versus weight of caja stones (*Spondias mombin* L.). All scales are in millimeters, except the length/width ratio

Table 7. Matrix of Pearson's linear correlation coefficients between pairs of variables: length, width, length/width ratio, thickness, and weight of the caja stones (*Spondias mombin* L.)

| Variables studied | Variables studied | | | | |
|-----------------------------|-------------------|-----------------------------|--------------------------|-----------------------|----------------------|
| | Length | Width | Length/width ratio | Thickness | Weight |
| Length | 1.00 | - | - | - | - |
| Width | 0.51 | 1.00 | - | - | - |
| Length/width ratio | 0.12 | -0.78 | 1.00 | - | - |
| Thickness | 0.54 | 0.86 | 0.60 | 1.00 | - |
| Weight | 0.02 | 0.05 | 0.07 | 0.06 | 1.00 |
| 95 % confidence intervals | Length × Width | Length × Width/Length/Width | Length × Thickness | Length × Weight | Width × Length/Width |
| | 0.42 to 0.59 | 0.01 to 0.23 | 0.46 to 0.62 | -0.10 to 0.13 | -0.82 to -0.73 |
| | Width × Thickness | Width × Weight | Length/Width × Thickness | Length/Width × Weight | Thickness × Weight |
| | 0.82 to 0.88 | -0.16 to 0.06 | -0.66 to -0.52 | -0.04 to 0.18 | -0.17 to 0.06 |
| P-value of Student's t-Test | <0.0001 | <0.0329 | <0.0001 | 0.7673 | <0.0001 |
| | <0.0001 | 0.39000 | <0.0001 | 0.2104 | 0.3457 |

The results of the descriptive statistical analysis (Table 1) show that the variables: length, width and thickness all in millimeters and the weight in grams of caja stones presented a high amplitude of variation, except the length/width ratio and low values for the relative variation of all of them measured by their coefficients of variation, and therefore optimal for the Pearson coefficients of variation, except the weight, since it presented a relatively high value for the coefficient of variation which corroborates that it is a characteristic (Sacramento & Souza, 2000), which shows a high degree of homogeneity of these evaluated characteristics, being the relative variation of length and weight different from the other variables, however there was similarity between thickness, width and length/width ratio. It was verified a reasonable degree of symmetry for all variables studied, that is, length, width, length/width ratio, thickness and weight of the caja stone, so they presented a symmetrical and mesocurtic distribution.

The results of the descriptive statistical analysis (Table 1) show that the variables: length, width and thickness all in millimeters and the weight in grams of caja stones presented a high amplitude of variation, except the length/width ratio and low values for the relative variation of all of them measured by their coefficients of variation, and therefore optimal for the Pearson coefficients of variation, except the weight, since it presented a relatively high value for the coefficient of variation which corroborates that it is a characteristic (Sacramento & Souza, 2000), which shows a high degree of homogeneity of these evaluated characteristics, being the relative variation of length and weight different from the other variables, however there was similarity between thickness, width and length/width ratio. It was verified a reasonable degree of symmetry for all variables studied, ie, length, width, length/width ratio, thickness and weight of the caja stone, so they presented a symmetrical and mesocurtic distribution.

It was observed a great similarity in the location or central tendency of the width and thickness of the caja stones, since the averages 16.0 and 16.5 mm, show that they are very close as far as the location of their distributions of values. On the other hand, the length, length/width ratio and weight of the stones showed a significant difference between the means whose results were 30.2 millimeters, 1.9 dimensionless, and 1.8 grams respectively. It is worth noting that averaging does not have to make real sense. However, it can be stated that the stones have a center of gravity in the distribution of these variables, which shows that the width and thickness of the stones are respectively close to 16.0 and 16.5 millimeters, or tend to vary around these values. On the other hand the weights are not very close to the average weight of 1.8 grams, since the value of the coefficient of variation has presented a value of 20.47%.

The relative variation between the variables measured by the coefficients of variation in this study was the most adequate for comparison of dispersion between samples, since they are independent of the magnitude of the variable, as well as of its unit of measurement, being, therefore, the most correct procedure. The values obtained from the coefficients of variation were 5.90%, 9.30%, 8.03% and 9.110% for length, width, length/width ratio and thickness, respectively, showing a low relative dispersion and similar or close to the four characteristics evaluated, thus revealing an extremely similar and statistically similar relative variability for these characteristics (Cramér, 1999; Beiguelman, 2002). The coefficient of variation for the weight of 20.47%, which shows a regular relative dispersion for this variable and thus shows a relatively moderate genetic variability that according to

Sacramento and Souza (2000), is inherent to this plant species (Ferreira, 2005; Oliveira et al., 2009; Zar, 2010; Fonseca & Martins, 2012).

Other results obtained, as can be seen in Table 1, were that the values obtained for the averages and medians were close to 30.2; 30.3, 16.0; 16.0, 1.9, 1.9, 16.5, 16.5 and 1.8; 1.8, respectively, which indicated a reasonable value for the degree of symmetry which were -0.4000; -0.1100; 0.1081; 0.0300 and 1.0000 for the length, width, length/width ratio, thickness and weight of the stones, respectively.

It is worth mentioning that in all symmetrical frequency distributions the mean or central arithmetic, median and fashion average position or trend measurements are the same or similar but the reciprocal is not true, the fact that the three means are close does not imply that the frequency distribution of their values are symmetrical (Gotelli & Ellison, 2011). On the other hand, the momentum coefficients of kurtosis were 0.4000; 0.3400; 0.1598; 1.7100 and 1.7700, respectively, showed that both the length, width, length/width ratio and thickness and weight presented a mesocurtic statistical distribution or series.

It is worth noting that the characterization of the degree of asymmetry and kurtosis of a distribution according to Spiegel and Stephens (2009), Zar (2010), Andrade and Ogliari (2010), Casella and Berger (2010); Gotelli and Ellison (2011), Fonseca and Martins (2012), can not be evaluated only the measures of position or central tendency as is the case of the mean and the median, but also through the coefficients of asymmetry and kurtosis, as well as through box plot diagrams, histograms and polygons of frequencies (Figures 1 to 10).

Asymmetry describes how the sample differs in the form of a symmetric distribution. A normal distribution has an asymmetry coefficient equal to zero. A distribution in which the value of the asymmetry coefficient is greater than zero has asymmetry to the right, that is, there is a long tail of larger observations, that is, to the right of the mean. In contrast if the asymmetry coefficient is less than zero it has asymmetry on the left, in this case there is a long tail of smaller observations, that is, to the left of the mean. Since kurtosis or flattening has its basis in the fourth central moment ($r = 4$), it measures the extent or peak at which the probability density is distributed in the tails versus the center of the distribution, ie in which the distribution is of heavy tail or light tail, when compared to a standard normal distribution (mesocurtic). Aggregate or platycurtic (flattened) distributions have a kurtosis coefficient less than zero, compared to the normal distribution, there is in this case more mass of probability at the center of the distribution and less probability in the tails. In contrast, leptokurtic (tapered) distributions have a value for the kurtosis coefficient greater than zero. Leptokurtic distributions have less mass of probability in the center and tails of relatively heavy probabilities (Gotelli & Ellison, 2011).

Regarding quantiles, the first quartile shows that 25% of the lowest values for the length, width and thickness of the stones reach a maximum of 29.1; 15.2; and 15.6 mm, respectively, and the dimensionless length/width ratio was 1.80. On the other hand the weight of the stones reaches this level in 1.6 grams. Already the 25% greater lengths, width and thicknesses of the caja stones, were represented by at least 31.5; 16.9 and 17.4 millimeters, respectively. However, the length/width ratio reaches a ratio of 2.0 and the weight of the stones in 2.1 grams, results corroborated by Figueiredo et al. (2007).

In this case the interquartile range obtained, which serves to verify the dispersion of the data in relation to the median and thus to identify the presence of atypical or Outlier's data, was 2.40; 1.75 and 1.87 mm, for the length, width and thickness respectively. The length/width ratio reached this level at a rate of 0.20, while the weight of the stones showed an amplitude value of 0.48 grams (Table 1).

The degree of association or simple linear correlation of Pearson whose physical definition field ranges from -1 to +1, including zero or null, between variables length and width, length and thickness, width and length/width, width and thickness and length/width ratio and thickness of caja kernels in millimeters, respectively showed values of magnitude 0.51; 0.54; -0.78; 0.86 and 0.60 (Table 7), which showed a direct or positive simple linear correlation between these characteristics, except for the width and length/width ratio, and thus these results strongly revealed three important aspects: direction, shape and strength or intensity of the association between the variables studied, and that was significant by Student's t-test hypothesis at 0.01% probability, showing, for example, that when the length value increases or decreases the width and thickness values also increase or decrease in a relatively close ratio, where the x and y pairs grow in the same proportion evidencing a normal bivariate distribution (Figure 11).

The other values obtained from Pearson's simple linear correlation coefficients showed that the other relationships between the other variables did not present a simple linear correlation statistically significant, which were 0.12; 0.02; 0.05; 0.07; and 0.06 (Table 7), for respectively length versus length to width ratio; length and weight; width and weight; (x, y) values do not present a bivariate normal distribution, nor at least

approximate, which means that the values of (x, y) reveals the lack of genetic studies of this plant species (Mardia, 1970; Ferreira, 2005; Figueiredo et al., 2011; Cecon et al., 2012; Costa, 2012).

Also with respect to the simple linear correlation coefficient of Pearson (r) obtained with the sample data, it was verified through Table 7 that one value stood out among the others ($r = 0.86$), and this value revealed or showed a relatively strong degree of simple linear association between the width and the thickness in millimeters of cashew kernels. This result is corroborated by the high sample size used and the value of the Student's t -test with low p -value or descriptive level of the test, in addition to the limits of the 95% confidence intervals, which proved to be a coefficient statistically different from zero, that is, it can not be considered as equal to a null value of the simple population correlation coefficient ρ (r_0), established under hypothesis and rejected with a low p value, which indicated a highly significant result from the point of statistical view.

Under the graphic, visual or geometric aspect, the dispersion diagram (Figure 11) showed a certain pattern of the cloud of pairs of points where we see an ascending direction from left to right, showing a positive coefficient, a shape close to one and also a relatively high association strength, thus, the longer the *caja* in millimeters, the denser and thicker the proportion, the denser and thicker the lump, the larger the lump will tend to be thicker also in millimeters. Exception was for the cloud of points between the width and the length/width ratio which showed a downward direction from left to right, ie, the more the width increases (or decreases) the length/width ratio decreases (or increases) proportional (Figure 11).

It is worth noting that a priori the pairs of variables length and width, length and length/width ratio, length and thickness and width and thickness of *cajarana* fruits all expressed in millimeters, except the length/width ratio that is adimensional, fit the distribution (Table 11), so that according to Mardia (1970), all conclusions obtained through statistical inference using either interval estimation or applying hypothesis or significance tests are guaranteed or assured by this presupposition.

As regards the application of the Z test to population mean μ (Table 1), very high values were observed for both length and width, and for length/width ratio, for thickness and also for weight of the *caja* stones, concluding that the mean values of these characteristics were highly significant, or statistically different from zero.

The box plot graph using the median and quartiles is a graphical tool widely used by researchers in general in the physical, biological, medical and social sciences, showing in the box the median, first and third quartiles. This graph also shows the lowest and highest score through the upper and lower boundary of vertical straight lines, which originate from the first and third quartiles, respectively. According to the results shown in Figures 1, 3, 5, 7 and 9, a strong concentration of data of length, width, length/width ratio, thickness of *caja* stones in millimeters and much less for weight in grams was observed, around the central value in the case the median or second quartile, which evidenced a high homogeneity of these observations, without acknowledging the presence of atypical values or outliers, supporting the value of the mean as an excellent summary measure of the central tendency of the data, thus becoming its current and subsequent analysis and interpretation, including the making of statistical inferences such as the construction of confidence intervals and the application of tests of hypotheses or of significance, as well as the adjustment of regression models for estimation and forecasting purposes, among others. Other types of analysis that have as such fact as assumption.

It is important to understand that the statistical analyzes of models with additive errors are based on the assumption of normality. The required normal distribution refers not to the variation of the data but to the residual variation between observations and the adjusted model. The systematic variation of the data is presumed to be the fixed effects of the models and the remainder of the random variation is due to small independent influences which produce waste with normal distribution (Bock, 1975).

In general, the results of the descriptive measures of location, variability, asymmetry and kurtosis can serve as a basis for future studies of descriptive analysis and statistical inference, for the comparison of different environments, genetic improvement studies, grouping of experiments in joint analysis, analysis of stability of cultivars, as well as in the construction of so-called components of variance and in the literature on the use of the term "biosynthesis" (Ferreira, 2005; Figueiredo et al., 2007; Spiegel & Stephens, 2009; Bussab & Morettin, 2010; Casella & Berger, 2010; Zar, 2010; Claudio & Stein, 2011; Cecon et al., 2012; Costa, 2012).

Given the results obtained previously, it is worth noting that the standard deviation and the variance are special cases of what the statisticians and physicists call the central momentum (MC). The central moment is the average of the deviation of all the observations in a set of data of the average of the observations, elevated to the power r . The first central moment ($r = 1$) measures the sum of the differences between each observation minus the sample mean (arithmetic), which is always equal to zero. The second central moment ($r = 2$) is the variance.

The third central moment ($r = 3$), divided by the standard deviation to the cube (s^3), is called asymmetry. Asymmetry describes how the sample differs in the form of a symmetric distribution. A normal distribution has an asymmetry coefficient equal to zero. A distribution in which the value of the asymmetry coefficient is greater than zero has asymmetry to the right, that is, there is a long tail of larger observations, that is, to the right of the mean. In contrast if the asymmetry coefficient is less than zero it has asymmetry on the left, in this case there is a long tail of smaller observations, that is, to the left of the mean. Since kurtosis or flattening has its basis in the fourth central moment ($r = 4$), it measures the extent or peak at which the probability density is distributed in the tails versus the center of the distribution, ie in which the distribution is of heavy tail or light tail, when compared to a standard normal distribution (mesocurtic).

Aggregate or platycurtic (flattened) distributions have a kurtosis coefficient less than zero, compared to the normal distribution, there is in this case more mass of probability at the center of the distribution and less probability in the tails. In contrast, leptokurtic (tapered) distributions have a value for the kurtosis coefficient greater than zero. Leptokurtic distributions have less mass of probability in the center and tails of relatively heavy probabilities (Gotelli & Ellison, 2011).

According to Gotelli and Ellison (2011), the law of large numbers proves that for an infinitely large number of observations, it is an approximation of the population mean, where is a sample of size n of a random variable Y with expected value (E) The standard deviation is simply the square root of the variance, so that is the same as the standard error of the mean, so we have an estimate of the standard deviation of the variance.

Depending on the inferential conclusions of the researcher, that is, if the conclusions, based on a single sample, are representative of the entire population, then the standard error of the mean should be used. If the conclusions are limited to the samples that the investigator has available, it is better to use the sample standard deviation. Large observational surveys covering large spatial scales with a significant number of a samples are likely to be representative of the population of interest as a whole, so the standard error of the mean should be used, while small, controlled experiments with few replicates are likely to be based on a single, and possibly unrepresentative, group of individuals, it is recommended to use the standard deviation to characterize or measure the degree of absolute dispersion of the samples.

5. Conclusions

The length, width, length/width ratio, thickness and weight of cashews presented a reasonable adjustment to the normal distribution of probabilities as well as a high total amplitude of variation, except the length/width ratio, as well as an optimum value for the coefficients of variation, except the weight, which showed a reasonable degree of homogeneity of these evaluated characteristics, including absolute and relative dispersion even though the values of the total amplitude being relatively high through the difference between the maximum and minimum values of the samples, and also by the values of the standard deviations and by the calculation of the coefficients of variation, the length and the thickness being the less dispersed variables than the width, length/width ratio and the weight of the stones.

The weight and width showed a strong adherence to the normal distribution of probabilities, evidenced by the histogram and frequency polygon as well as through the result of D'Agostino-Pearson's normality test. On the other hand, length, length/width ratio and thickness fit roughly the normal or Gaussian distribution of probabilities.

A regular degree of symmetry was found for the frequency distributions of length, width, length/width ratio, thickness as well as the weight of cashew kernels.

A degree of flatness or kurtosis of the mesocurtic type was found for the series of values of length, width, length/width ratio, thickness and weight of the stones.

The characteristics of length and width, length and width, width and length/width, width and thickness, as well as length/width ratio and thickness of caja stones showed significant linear correlation of Pearson significant and mean values with highly significant statistical differences, being the stronger and direct relationship between width and thickness.

The use of Pearson's linear simple correlation coefficients was validated considering that the univariate data referring to the characteristics of the caja evaluated in this work, adjusted to the theoretical distribution of normal or Gaussian probability, at least approximately.

References

- Andrade, D. F., & Ogliari, P. J. (2010). *Estatística Para as Ciências Agrárias e Biológicas Com Noções de Experimentação* (2nd ed.). Editora da UFSC.
- Azevedo, D. M., Mendes, A. M., & Figueredo, A. F. (2004). Característica da germinação e morfologia do endocarpo e plântula de taperebá (*Spondias mombin* L.)-Anarcadiaceae. *Revista Brasileira de Fruticultura*, 26, 534-537. <https://doi.org/10.1590/S0100-29452004000300038>
- Baskin, C. C., & Baskin, J. M. (1998). *Seeds: Ecology, biogeography and evolution of dormancy and germination* (p. 666). London: Academic Press.
- Beiguelman, B. (2002). *Curso prático de bioestatística* (5th ed., p. 274). Ribeirão Preto: FUNPEC.
- Bock, R. D. (1975). *Multivariate statistical methods in behavioral research* (p. 623). McGraw Hill, Universidade de Minnesota.
- Bosco, J., Soares, K. T., Aguiar Filho, S. P., & Barros, R. V. (2000). *A cultura da cajazeira* (p. 29). EMEPA-PB.
- Bussab, W. O., & Morettin, P. A. (2010). *Estatística básica* (6th ed., p. 540). São Paulo: Saraiva.
- Carmo Filho, F., Espíndola Sobrinho, J., & Maia Neto, J. M. (1991). Dados meteorológicos de Mossoró (Janeiro de 1989 a Dezembro de 1990). *Coleção Mossoroense, Série C, 630* (p. 110). Mossoró: ESAM.
- Carpanezzi, A. A., & Marques, L. C. T. (1981). Germinação de sementes de jutaí-açu (*Hymenaea courbaril* L.) e de jutaí-mirim (*H. parvifolia* Huber) escarificadas com ácido sulfúrico comercial. *Circular Técnica 19* (p. 15). EMBRAPA-CPATU.
- Carvalho, J. E. U., Nascimento, W. M. O., & Muller, C. H. (1998). Características físicas e de germinação de sementes de espécies frutíferas nativas da Amazônia. *Boletim de Pesquisa 203* (p. 18). EMBRAPA-CPATU.
- Casella, G., & Berger, R. L. (2010). *Inferência estatística* (p. 612). São Paulo: Editora Cengage.
- Cecon, P. R., Silva, A. R., Nascimento, M., & Ferreira, A. (2012). *Métodos estatísticos* (p. 229). Editora UFV, Brazil.
- Claudio, L. C., & Stein, C. E. (2011). *Estatística descritiva e teoria das probabilidades* (2nd ed., p. 213) Edifurb.
- Costa, G. G. O. (2012). *Curso de estatística inferencial e probabilidades: Teoria e Prática* (p. 370). Editora Atlas.
- Cramér, H. (1999). *Mathematical methods of Statistics* (p. 575). Princeton University Press, USA.
- Feitosa, S. S. (2007). *Nutrição mineral e adubação da cajazeira (Spondias mombin L.) na Zona da Mata Paraibana* (p. 50, Dissertação (Solos e Nutrição de Plantas), Areia, PB).
- Fenner, M. (1993). *Seed ecology* (p. 151). Chapman & Hall.
- Ferreira, D. F. (2005). *Estatística básica* (p. 625). UFLA, Brazil.
- Figueiredo, F., Figueiredo, A., Ramos, A., & Teles, P. (2007). *Estatística descritiva e probabilidades-problemas resolvidos e propostos com aplicações em R* (p. 420). Escolar Editora, Lisboa, Portugal.
- Fonseca, J. S., & Martins, G. A. (2012). *Curso de estatística* (6th ed., p. 320). Atlas.
- Gotelli, N. J., & Ellison, A. M. (2010). *Princípios de estatística em ecologia* (p. 532). Artmed Editora.
- Gusmão, E., Vieira, F. A., & Júnior, E. M. F. (2006). Biometria de frutos e endocarpos de murici (*Byrsonima verbascifolia* Rich. Ex. A. Juss). *Revista Cerne*, 12, 84-91.
- Kronka, S. N., & Banzato, D. A. (1995). *(Estat) system for statistical analysis 2* (3rd ed., p. 243). Jaboticabal: Funep.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530. <https://doi.org/10.1093/biomet/57.3.519>
- Mathus, M. T., & Lopes, J. C. (2007). Morfologia de frutos e plântulas e germinação de sementes de *Erythrina variegata* L. *Revista Brasileira de Sementes*, 29, 8-15. <https://doi.org/10.1590/S0101-31222007000300002>
- Mendonça, A. R. U., Moura, C. F. H., Alves, R. E., Figueiredo, R. W., & Souza, V. A. B. (2008). *Caracterização física de frutos da cajazeira (Spondias mombin L.) oriundos da região meio-norte do Brasil*. XX Congresso Brasileiro de Fruticultura 54th Annual Meeting of the Interamerican Society for Tropical Horticulture. 12 a 17 de Outubro de 2008, Centro de Convenções, Vitória, ES.

- Oliveira, M. S., Bearzoti, E., Vilas Boas, F. L., Nogueira, D. A., Nicolau, L. A., & Oliveira, H. S. S. (2009). *Introdução à estatística* (p. 329). Lavras, MG: Editora da Universidade Federal de Lavras.
- Pereira, K. S. N. (1996). *Cajá (Spondias mombin L.): Características da unidade de dispersão, propagação sexuada e assexuada* (p. 31). Areia: UFPB.
- Quadros, B. R., Ribeiro, O. D., Rodrigues Júnior, O. M., Nascimento, W. M. O., Cavariani, C., & Costa, E. N. (2015). Biometria do endocarpo de taperebá (*Spondias mombin* L.-Anacardiaceae). *Revista Cultivando o Saber*, 8, 250-256.
- R Version 3.1.1. (2018). Viena, Áustria: Foundation for Statistical Computing.
- Sacramento, C. K., & Souza, F. X. (2000). *Cajá (Spondias mombin L.)* (Série Frutas Nativas, p. 42). Jaboticabal: Funep.
- Sacramento, C. K., & Souza, F. X. (2009). Cajá. In J. A. Santos-Serejo, J. L. L. Dantas, C. V. Sampaio, & Y. S. Coelho (Eds.), *Fruticultura tropical: Espécies regionais e exóticas* (p. 506). Brasília, DF: Embrapa Informação Tecnológica.
- Souza, A. A., Bruno, R. L. A., Lopes, K. P., Cardoso, G. D., Pereira, W. E., & Cazé Filho, J. (2005). Semillas de *Spondias tuberosa* oriundos de frutos cosechados en cuatro estadios de maduración y almacenadas. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 9, 372-378. <https://doi.org/10.1590/S1415-43662005000300012>
- Spiegel, M. R., & Stephens, L. J. (2009). *Estatística* (4th ed., p. 597). Bookman.
- Zar, J. H. (2010). *Biostatistical analysis* (5th ed., p. 944). Prentice Hall.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).